

The Concept of Major Depression

III. Concurrent Validity of Six Competing Operational Definitions for the Clinical ICD-9 Diagnosis

Michael Philipp, Wolfgang Maier, and Cynthia D. Delmo

Department of Psychiatry, University of Mainz, Untere Zahlbacher Strasse 8, W-6500 Mainz, Federal Republic of Germany

Received June 14, 1990

Summary. The comparative validity of six operational diagnoses of major depression was evaluated in 600 psychiatric inpatients using the independently assessed clinical ICD-9 diagnoses as a yardstick. Agreement with, and positive predictive value for the ICD-9 categories of pure (endogenous and psychogenic) depression served as validation criteria; sensitivity of major depression diagnoses for detecting ICD-9 bipolar depressions was additionally used for examining the adequacy of width, time and exclusion criteria of the competing operational definitions. Three essential results were found. First, the “old” diagnostic definitions of RDC and FDC are superior to all newer definitions because they define the time criteria and the schizophrenic exclusion criteria more adequately than, for example, both DSM-III and DSM-III-R definition. Secondly, the current ICD-10 definition of 1989 (“mild”, “moderate” or “severe” depression) comes closer to the concurrent validity of RDC and FDC than DSM-III, DSM-III-R and the previous ICD-10 definition of 1987. Thirdly, using the criterion of identifying a high proportion of ICD-9 bipolar depressions, all six competing diagnostic systems are too restrictive. Evaluations of predictive and criterion-related validity will be needed to substantiate these findings.

Key words: Major depression – Operational diagnosis – Concurrent validity – ICD-10 – DSM-III-R

Introduction

The empirical evaluation of the agreement between six operational definitions of major depression (Philipp et al. 1990b) revealed that minor differences of diagnostic algorithms, especially of defining the obligatory mood criterion, the syndrome cut-off, the time criteria and the exclusion criteria, led to major discrepancies in both diagnostic base rates and pairwise overlap between diag-

noses. Yet the assumption of a preceding descriptive comparison of the diagnostic algorithms (Philipp et al. 1990a) could be confirmed: the current ICD-10 definition of “mild”, “moderate” or “severe” depression (WHO 1989) refers to the same concept of major depression operationalized in DSM-III, DSM-III-R, RDC and FDC.

The conclusions of this first evaluation of agreement between diagnoses are limited by the absence of an independent validation criterion. It is the aim of this paper to present initial comparative results concerning the concurrent validity of the competing operational definitions. The independently assessed clinical ICD-9 diagnoses are used in order to evaluate the agreement and positive predictive value of major depression definitions for the detection of ICD-9 pure depression (i.e. endogenous or psychogenic depression); furthermore, ICD-9 bipolar depression diagnoses are used for examination of the adequacy of width, time and exclusion criteria of the competing operational diagnoses.

Evaluations are done in the same sample of 600 psychiatric inpatients for examination of the agreement between diagnoses. The description of the patients and the methods used is therefore restricted to additional aspects that were not described in the previous paper (Philipp et al. 1990b).

Methods

Patients and Assessment of Diagnoses. Six hundred psychiatric inpatients were included in the study applying the Polydiagnostic Interview (PODI; Philipp and Maier 1986) within the first 4 weeks after admission. Details of the interview procedure and the computerized calculation of operational diagnoses (Delmo 1988; Delmo and Philipp 1989) were described in a previous paper in this series (Philipp et al. 1990b). Clinical diagnoses were established according to ICD-9 at discharge by the treating psychiatrist who was blind to the diagnostic results of the PODI. The PODI-assessed operational diagnoses were not completely independent of the clinical diagnosis. Within the PODI assessment the research assistants discussed the codings of PODI items with the treating psychiatrist, receiving additional clinical information and integrating the

results of their discussion in the final codings of the PODI. This partial dependence seemed acceptable, since the objectivity and validity of the PODI diagnoses were increased without interfering with the main aim of the study, the comparison of relationships between agreement coefficients.

Statistical Calculations. Six concurrent definitions of major depression were studied: Feighner Diagnostic Criteria (FDC; Feighner et al. 1972; Feighner 1981), Research Diagnostic Criteria (RDC; Spitzer et al. 1978), Diagnostic and Statistical Manual of Mental Disorders, third edition (DSM; American Psychiatric Association 1980), Diagnostic and Statistical Manual of Mental Disorders, third edition, revised (DSR; American Psychiatric Association 1987), ICD-10 Diagnostic Criteria for Research, draft April 1987 (I87; WHO 1987) and ICD-10 Diagnostic Criteria for Research, draft April 1989 (I89; WHO 1989). Additionally, two new definitions (one wide and one narrow) of major depression were studied, the wide one being defined as meeting at least one of the above-mentioned six operational definitions and the narrow one being defined as meeting all of them.

For each of these eight definitions the distributions of patients meeting the definition of major depression were calculated for all main diagnostic categories of ICD-9 in general and for all ICD-9 depression diagnoses in particular. Coefficients of agreement corrected by chance (kappa; Cohen 1960) were calculated between each definition of major depression and the following groups of ICD-9 diagnoses: (a) any depression (either of ICD 296.1, 296.3, 296.4, 300.4, 301.1, 308.0, 309.0, 309.1, or 311); (b) bipolar depression (ICD 296.3 or 296.4).

The amount of agreement was evaluated for every level of the operational definition process (obligatory mood inclusion criterion, pure syndrome definition, syndrome with time criterion, full diagnosis). For better comparability, certain exclusion criteria of the FDC (history of nonaffective psychiatric disorders) and both ICD-10 diagnoses I87 and I89 (bipolar depression) were not applied at the level of full diagnosis. Sensitivity, specificity and positive predictive value (ppv50) were completed for each diagnosis at each level of the operational definition. Finally, the level of agreement between the wide and the narrow definition and the clinical ICD-9 diagnosis of depression was calculated in order to evaluate the influence of restrictive and nonrestrictive definitions of major depression on the agreement with the ICD-9 diagnosis.

Results

The overall distribution of ICD-9 diagnostic main categories in the sample of 600 inpatients (Table 1) shows a clear-cut dominance of affective psychoses (ICD 296: 40.0%), followed by schizophrenic psychoses (ICD 295:

28.8%) and neuroses (ICD 300: 12.0%). All other diagnostic main categories contain less than 10% of patients.

Within the subsamples of major depression, the frequency of patients is lower for schizophrenic psychoses (ICD 295) and for alcohol dependence and abuse (ICD 303); it is higher for affective psychoses (ICD 296) and for adjustment disorders (ICD 309) (Table 1). Between the six operational diagnoses of major depression, large differences are seen in the corresponding frequencies of schizoaffective psychoses were RDC (7.5%) and FDC (9.9%) show rather low percentages while the other diagnoses present nearly double the percentage (DSR: 17.7%; DSM: 17.2%; I87: 16.8%; I89: 15.3%) (Table 1). In all other diagnostic main categories, the differences between the six major depression definitions are smaller; for unipolar and bipolar endogenous depression, in particular, the range is only 6.7% (DSR: 55.2%; RDC: 61.9%); for neurotic depression the range is only 2.3% (DSM: 11.5%; RDC: 13.8%). Both diagnostic main categories FDC and RDC have the largest percentage of patients in comparison with the other diagnostic systems.

Within those ICD-9 categories comprising depressive syndromes (Table 2) schizoaffective depressions again are lowest in RDC (4.5%) and FDC (5.7%); the highest percentage of schizoaffective depressions is found in I87 (8.6%). Unipolar endogenous depression (ICD 296.1) has the highest frequencies in RDC (51.5%), FDC (50.2%) and I89 (49.5%); the other three diagnostic systems all shown 46.4% unipolar endogenous depressions. For bipolar depression (ICD 296.3 + 4), RDC again is highest (7.5%) and DSR lowest (5.8%). The highest percentage of neurotic depressions is found in the FDC (6.7%), the lowest in DSM (3.9%); all other diagnostic systems range between 4.0% and 4.5%. For adjustment disorders, depressed type (309.0 + 1) the range is very narrow (between 6.4% in FDC and 7.1% in RDC and DSM) (Table 2).

The percentage of all ICD-9 depression diagnoses taken together is higher in the major depression groups than in the total sample (Table 3). Yet it differs between groups: schizoaffective depression (ICD 295.7) is diagnosed about four times more often in the total group (22.8%) than in any of the major depression groups (range between 4.5% and 8.6%); the frequency of en-

Table 1. Distribution of clinical ICD-9 diagnoses in 600 psychiatric inpatients

	Total number	Percentage of patients with clinical ICD-9 main diagnoses														
		290-294	295	296	297	298	300	301	303	304	305	307	308	309	311	Others
Total group	<i>n</i> = 600	1.0%	28.8%	40.0%	0.3%	0.7%	12.0%	2.0%	7.0%	0.5%	0.5%	0.2%	0.8%	5.5%	0.2%	0.5%
<i>Major Depression</i>																
FDC	<i>n</i> = 283	1.1%	9.9%	59.4%	0%	0.4%	13.4%	3.2%	1.8%	0.4%	0.7%	0.4%	1.4%	7.1%	0.4%	0.7%
RDC	<i>n</i> = 268	0.8%	7.5%	61.9%	0%	0.4%	13.8%	3.0%	2.2%	0.4%	0.4%	0.4%	1.1%	7.1%	0.4%	0.7%
DSM	<i>n</i> = 279	1.1%	17.2%	55.6%	0%	0.4%	11.5%	2.2%	1.1%	0.7%	0.4%	0.4%	0.7%	8.2%	0.4%	0.4%
DSR	<i>n</i> = 277	1.1%	17.7%	55.2%	0%	0.4%	11.9%	2.2%	1.1%	0.4%	0.4%	0.4%	0.7%	7.9%	0.4%	0.4%
I87	<i>n</i> = 232	1.3%	16.8%	56.5%	0%	0.4%	12.5%	2.6%	0.9%	0%	0.4%	0.4%	0.4%	6.9%	0.4%	0.4%
I89	<i>n</i> = 275	0.7%	15.3%	58.2%	0%	0.4%	11.6%	2.6%	1.1%	0.4%	0.4%	0.4%	1.1%	7.3%	0.4%	0.7%

Table 2. Distribution of clinical ICD-9 depression diagnoses in 600 psychiatric inpatients

	Percentage of patients with clinical ICD-9 depression diagnoses							
	2957	296 unp.	296 bip.	296 NOS	3004	3011	3080	3091 + 2
Total sample	22.8%	29.3%	4.5%	1.0%	3.2%	0.2%	0.8%	4.8%
FDC	5.7%	50.2%	6.7%	2.5%	6.7%	0.4%	1.4%	6.4%
RDC	4.5%	51.5%	7.5%	3.0%	4.5%	0%	1.1%	7.1%
DSM	6.8%	46.6%	6.1%	2.9%	3.9%	0%	0.7%	7.1%
DSR	7.2%	46.6%	5.8%	2.9%	4.3%	0%	0.7%	6.8%
I87	8.6%	46.6%	7.3%	2.6%	4.3%	0.4%	0.4%	6.5%
I89	7.3%	49.5%	6.6%	2.2%	4.0%	0%	1.1%	6.6%

Table 3. Frequency of main clinical ICD-9 depression categories in 600 psychiatric inpatients

	Patients with ICD-9 endogenous depression (296.1, 296.3, 296.4)	Patients with ICD-9 psychogenic depression (300.4, 301.1, 308.0, 309.0, 311)	Patients with ICD-9 schizo-affective depression (295.7)	Patients with any ICD-9 depression
Total sample	33.8%	9.2%	22.8%	65.8%
FDC	56.9%	13.8%	5.7%	80.4%
RDC	59.0%	13.1%	4.5%	79.6%
DSM	52.7%	12.2%	6.8%	74.5%
DSR	52.4%	12.3%	7.2%	74.7%
I87	53.9%	12.1%	8.6%	77.1%
I89	56.1%	12.0%	7.3%	77.7%

Table 4. Quantitative relation of main clinical ICD-9 depression categories in 600 psychiatric inpatients

	Ratio unipolar: bipolar	Ratio endogenous: psychogenic	Ratio pure depression: schizo-affective depression	Ratio any depression: no depression
Total sample	6.5	3.7	1.9	1.9
FDC	7.5	4.1	13.1	4.1
RDC	6.9	4.5	16.7	3.9
DSM	7.6	4.3	10.0	2.9
DSR	8.0	4.3	9.4	3.0
I87	6.4	4.5	8.0	3.4
I89	7.5	4.7	9.6	3.4

ogenous and psychogenic depression on the other hand is approximately 50% higher in the major depression groups than in the total group (Table 3). Comparing the competing major depression definitions the highest amount of any ICD-9 depression is found for the FDC (80.4%) and the RDC (79.6%) and the lowest for DSM (74.5%) and DSR (74.7%). The range of ICD-9 endogenous depression is rather narrow with RDC being high-

est (59.0%) and DSR lowest (52.4%); this range is even smaller for ICD-9 psychogenic depression, where the percentage of FDC is highest (13.8%) and that of I89 is lowest (12.0%). The ICD-9 diagnosis of schizoaffective depression is found nearly twice as often in the I87 group (8.6%) as in the RDC Group (4.5%); RDC (4.5%) and FDC (5.7%) have the lowest percentage of ICD-9 schizoaffective depression diagnoses, while I87 (8.6%) and I89 (7.3%) have the highest percentage (Table 3).

Table 4 shows the quantitative relationships between main ICD-9 depression groups. The ratio of ICD-9 unipolar and bipolar endogenous depression in the total group (6.5:1) is within the range of all major depression subgroups: the lowest ratio (7.5:1) is found for I87, the highest (8.0:1) for DSR. The ratio of endogenous and psychogenic depression is lower in the total group (3.7:1) than in the major depression subgroups, where the range is quite narrow (lowest for FDC, 4.1:1; highest for I89, 4.7:1). The ratio of pure depression (i.e. endogenous or psychogenic) and schizoaffective depression differs extremely between the total group (1.9:1) and the major depression subgroups (between 8.0 and 16.7:1). Accordingly, the wide:RDC shows the highest ratio (16.7:1), followed by the FDC (13.1:1); I87 has the lowest ratio (8.0:1) with DSM (10.0:1), I89 (9.6:1) and DSR (9.4:1) being very close (Table 4). The ratio of any ICD-9 depression and no ICD-9 depression is also clearly higher in the major depression samples (DSM 2.9:1 to FDC 4.1:1) than in the total sample (1.9:1).

Agreement corrected by chance (kappa) between the operational definitions of major depression and the ICD-9 diagnosis of depression is rather low at the level of the obligatory mood inclusion criterion (kappa between 0.35/FDC and 0.42/I89) and at the level of pure syndrome definition (kappa between 0.38/RDC/DSM, and 0.41/I89); the introduction of time criteria rather lowers chance corrected agreement (kappa between 0.33/I87 and 0.38/FDC) (Table 5). The additional application of exclusion criteria increases kappa values; kappa is now highest for FDC and RDC (each 0.53) and lowest for I87 (0.37). The relationship between sensitivity and specificity changes throughout the operational definition process: at the level of the obligatory mood inclusion criterion sensitivity is highest (between 95.0% and 98.5%) and specificity is lowest (between 39.8% and 50.3%); this difference vanishes at the level of full diagnosis: sensitivity and specificity are nearly equivalent for all diag-

Table 5. Agreement of six definitions of major depression with the clinical ICD-9 diagnosis of any pure depression (endogenous: ICD 296.1, 296.3, 296.4, or psychogenic: ICD 300.4, 301.1, 308.0, 309.0, 309.1, 311)

Dia- gnostic system	Agreement ICD-9 diagnosis (kappa)	Sensitivity	Specificity	Positive predictive value (50)
<i>Obligatory mood inclusion criterion</i>				
FDC	0.35	98.5	39.8	69.1
RDC	0.36	98.5	41.2	69.8
DSM	0.37	98.5	41.5	70.0
DSR	0.36	98.5	40.9	69.7
I87	0.40	95.0	48.5	71.8
I89	0.42	95.0	50.3	72.6
<i>Pure syndrome definition</i>				
FDC	0.39	88.4	52.6	70.5
RDC	0.38	88.4	52.3	70.4
DSM	0.38	93.0	47.4	70.2
DSR	0.37	93.0	47.1	70.1
I87	0.40	87.2	54.7	70.9
I89	0.41	93.4	50.6	72.0
<i>Syndrome with time criterion</i>				
FDC	0.38	81.8	57.9	69.8
RDC	0.36	81.0	57.0	69.0
DSM	0.36	73.3	64.0	68.7
DSR	0.35	72.5	63.7	68.1
I87	0.33	59.7	73.4	66.5
I89	0.36	77.5	60.2	68.9
<i>Full diagnosis</i>				
FDC ^a	0.53	77.5	75.7	76.6
RDC	0.53	74.8	78.1	76.5
DSM	0.41	70.2	71.4	70.8
DSR	0.40	69.4	71.4	70.4
I87 ^a	0.37	59.3	76.9	68.1
I89 ^a	0.46	72.5	74.3	73.4

^a For FDC, I87 and I89 the additional exclusion of secondary and bipolar patients was omitted

noses with the exception of I87, where sensitivity (59.3%) is much lower than specificity (76.9%) (Table 5).

ppv50 is rather low in the first three steps of the operational definition process: it varies between 69.1% and 72.6% at the first level, between 70.1% and 72.0% at the second level, and between 66.5% and 69.8% at the third level (Table 5). Again the introduction of time criteria rather reduces ppv50, while the application of exclusion criteria leads to an increase again. Now FDC (76.6%) and RDC (76.5%) are highest, DSM (70.8%) and DSR (70.4%) are intermediate, and I87 (68.1%) and I89 (73.4%) are most discrepant.

Table 6 displays the sensitivity of operational diagnoses of major depression for the detection of cases clinically classified as bipolar endogenous depression. The comparison of obligatory mood inclusion criterion and pure syndrome definition does not show any significant difference; only I87 drops in sensitivity from 92.6% to 88.9%. By the introduction of time criteria the sensitiv-

Table 6. Sensitivity of six definitions of major depression for the clinical ICD-9 diagnosis of bipolar depression (ICD 296.3 + 4)

Dia- gnostic system	Obligatory mood criterion	Pure syndrome definition	Syndrome with time criterion	Full diagnosis
FDC ^a	96.3%	92.6%	81.5%	70.3%
RDC	96.3%	96.3%	88.9%	74.1%
DSM	96.3%	96.3%	74.1%	63.0%
DSR	96.3%	96.3%	70.4%	59.3%
I87 ^a	92.6%	88.9%	63.0%	63.0%
I89 ^a	96.3%	96.3%	81.5%	66.7%

^a For FDC, I87 and I89 the additional exclusion of secondary and bipolar patients was omitted

ity for all diagnoses is reduced even further. Yet this drop varies considerably: it is highest for I87 (88.9% to 63.0%) and for both DSM-III diagnoses (DSM: 96.3% to 74.1%; DSR: 96.3% to 70.4%); the drop is lowest in RDC (from 96.3% to 88.9%) and moderate in FDC (from 92.6% to 81.5%) and I89 (from 96.3% to 81.5%) (Table 6). In the last step of applying exclusion criteria I87 is not reduced any more in sensitivity; all other diagnoses are reduced by at least 10%. At this level of full diagnosis RDC remains the most sensitive (74.1%), followed by FDC (70.3%) and I89 (66.7%); DSR is lowest with a sensitivity of only 59.3%; DSM and I87 are only a little higher, 63.0% each.

Applying a wide definition of major depression (at least one of the six operational diagnoses must be met) sensitivity for any ICD-9 pure depression remains high (> 91%) throughout the four steps of the operational definition process with specificity increasing from 39.8% for the obligatory mood inclusion criterion to 59.7% for the full diagnosis (Table 7). Agreement corrected by chance (kappa = 0.48) and ppv50 (= 75.4%) are of medium level compared to the coefficients of the single operational definitions. The narrow definition (all six operational diagnoses have to be met) is lower in agreement (kappa = 0.40) and predictive value (ppv 50 = 69.3%) than the wide definition on the level of full diagnosis; sensitivity and specificity are inverse at this level in comparison with the wide definition (Table 7).

The sensitivity of the wide definition of major depression for detecting ICD-9 bipolar depression is 88.9% at the level of full diagnosis (Table 8); this is much higher than for any single operational definition, where RDC is highest with a sensitivity of 74.1%. On the other hand, the sensitivity of the narrow definition is much lower (48.2%) than for the lowest single operational definition (DSM and I87: 63.0%).

Discussion

Distribution of ICD-9 Diagnoses

The distribution of ICD-9 diagnoses in the total sample under study is rather untypical of psychiatric inpatient departments with predominantly acute psychiatric cases: affective psychoses (ICD 296) predominate, whereas in

Table 7. Agreement between wide and narrow definition of major depression and the clinical ICD-9 diagnosis of any pure depression

Wide definition of major depression								
Mood criterion			Pure syndrome		Syndrome and time		Full diagnosis ^a	
<div>- +</div>			<div>- +</div>		<div>- +</div>		<div>- +</div>	
ICD-9 any depression	- 136	206	- 157	185	- 169	173	- 204	138
	+ 3	255	+ 11	247	+ 19	239	+ 23	235
	Kappa = 0.35		Kappa = 0.39		Kappa = 0.39		Kappa = 0.48	
	Sens. = 98.8%		Sens. = 95.7%		Sens. = 92.6%		Sens. = 91.1%	
	Spec. = 39.8%		Spec. = 45.9%		Spec. = 49.4%		Spec. = 59.7%	
	ppw50 = 69.3%		ppw50 = 70.8%		ppw50 = 71.0%		ppw50 = 75.4%	
Narrow definition of major depression								
Mood criterion			Pure syndrome		Syndrome and time		Full diagnosis ^a	
<div>- +</div>			<div>- +</div>		<div>- +</div>		<div>- +</div>	
ICD-9 any depression	- 183	159	- 199	143	- 258	84	- 299	43
	+ 18	240	+ 40	218	+ 113	145	+ 126	132
	Kappa = 0.44		Kappa = 0.41		Kappa = 0.32		Kappa = 0.40	
	Sens. = 93.0%		Sens. = 84.5%		Sens. = 56.2%		Sens. = 51.2%	
	Spec. = 53.5%		Spec. = 58.2%		Spec. = 75.4%		Spec. = 87.4%	
	ppw50 = 73.3%		ppw50 = 71.3%		ppw50 = 65.8%		ppw50 = 69.3%	

^a For FDC, I87 and I89 the additional exclusion of secondary and bipolar patients was omitted

Table 8. Sensitivity of wide and narrow definition of major depression and the clinical ICD-9 diagnosis of bipolar depression

Wide definition of major depression								
	Mood criterion		Pure syndrome		Syndrome and time		Full diagnosis ^a	
	-	+	-	+	-	+	-	+
ICD-9	-138	435	-167	406	-187	386	-224	349
bipolar depression	+ 1	26	+ 3	24	+ 1	26	+ 3	24
	Sens. = 96.3%		Sens. = 96.3%		Sens. = 96.3%		Sens. = 88.9%	
Narrow definition of major depression								
	Mood criterion		Pure syndrome		Syndrome and time		Full diagnosis ^a	
	-	+	-	+	-	+	-	+
ICD-9	-199	374	-236	337	-359	214	-411	162
bipolar depression	+ 2	25	+ 3	24	+ 12	15	+ 14	13
	Sens. = 92.6%		Sens. = 88.9%		Sens. = 55.6%		Sens. = 48.2%	

^a For FDC, I87 and I89 the additional exclusion of secondary and bipolar patients was omitted

other institutions schizophrenic psychoses (ICD 295) are most frequent. This feature has been characteristic of the patient structure of this department for many years; it does not reflect hidden selection mechanisms. It seems to be more importance that in the total sample as well as in the subsamples of major depression endogenous depression (ICD 296.1,3,4) is diagnosed four times more often than any type of psychogenic depression (ICD 300.4, 301.1, 308.0, 309.0,1, 311). A ratio of 6.5:1 between unipolar and bipolar endogenous depression in the total group is quite unusual compared with a ratio of about 3:1 normally reported for psychiatric inpatient samples. Bipolar endogenous depressions are diagnosed in nearly 5% of all patients; this seems to correspond to the frequency in other psychiatric inpatient departments;

we can therefore conclude that unipolar endogenous depressions are overrepresented by about 30%. We have no indication that patients with unipolar endogenous depression have been positively overselected; we therefore assume that our clinical diagnostic habits differ from those of other departments in overestimating endogenous depression and underestimating psychogenic depression.

Yet this overestimation of endogenous depression does not exclude the use of the combined endogenous and psychogenic depression group as a yardstick for comparing sensitivity, specificity and ppw50 of the six competing operational diagnoses of major depression. Previous studies (Philipp and Maier 1985, 1987; Maier et al. 1986; Hiller et al. 1988) have shown that operational

diagnoses of major depression incorporate not only patients clinically diagnosed as endogenous depression (ICD 196) but also a relevant proportion of patients clinically diagnosed as neurotic depression (ICD 300.4) or depressive adjustment reaction (ICD 309.0,1). The adequate yardstick for evaluation of the concurrent validity of operational definitions of major depression therefore is the combined group of ICD-9 endogenous and psychogenic depression. Additionally the adequate frequency of ICD-9 diagnoses of bipolar endogenous depression justifies using the sensitivity of major depression in the detection of this diagnostic class for judging the adequacy of the width of the competing operational definitions. Finally, the percentage of patients with ICD-9 diagnoses of schizophrenic psychoses is high enough to examine the adequacy of exclusion criteria by using the exclusion of patients with schizophrenic psychoses (ICD 295) as a yardstick.

Concurrent Validity for ICD-9 Pure Depression

At the level of full diagnosis FDC and RDC are clearly higher in both aspects of concurrent validity (agreement and positive predictive value) than the other four diagnostic systems when using the ICD-9 diagnostic groups of pure depression (endogenous and psychogenic) as a validation criterion. The current ICD-10 definition I89 follows in third place, whereas the previous ICD-10 definition I87 is lowest in concurrent validity.

The stepwise conducted comparison allows us to see that the superiority of FDC, RDC and I89 in concurrent validity is exclusively due to the more intensive exclusion of patients with concurrent schizophrenic symptoms: it is only at the level of full diagnosis where this superiority can be seen; in the three preceding steps of the operational definition process there are no significant differences in agreement and ppv50. Consequently we can conclude that the differences in defining the obligatory mood inclusion criterion, the syndrome cut-off and the time criterion are of no relevance for the concurrent validity with the independently assessed clinical ICD-9 diagnosis of pure depression.

Adequacy of Exclusion Criteria

The exclusion criteria applied at the level of full diagnoses were restricted to schizophrenic symptoms; all patients had functional psychiatric disorders, i.e. somatic disorders had been excluded from the study; furthermore the exclusion criteria referring to psychiatric history (previous manic or nonaffective episodes) were not applied. The superiority of FDC and RDC – and in part of I89 – in concurrent validity must therefore reflect a greater adequacy in excluding patients with concomitant schizophrenic symptoms. The descriptive comparison of exclusion criteria (Philipp et al. 1990a) has shown that FDC, RDC and I89 exclude patients from the diagnosis of major depression, when schizophrenic symptoms occur simultaneously with the depressive syndrome; in contrast to this both DSM-III definitions and I87 restrict the exclusive power of schizophrenic symptoms to those patients

in whom these symptoms are also present in the absence of depressive symptoms. Using the concurrent validity with the ICD-9 diagnosis as an independent criterion our data clearly demonstrate that both the DSM-III and the I87 definition of schizophrenic exclusion criteria are too restrictive; FDC, RDC and I89 are more adequate in ruling out the diagnosis of major depression if schizophrenic symptoms occur simultaneously with the depressive syndrome. It is apparent that the current ICD-10 definition I89 defines the schizophrenic exclusion criterion more adequately than DSM-III-R and the previous ICD-10 definition I87.

Sensitivity for ICD-9 Bipolar Depression

In this study the sensitivity of major depression definitions for detecting patients with ICD-9 bipolar depression diagnoses served as a measure of the adequacy of the width of the different diagnostic algorithms. Bipolar depression showed the most valid diagnostic category of all depression diagnoses; the clinical ICD-9 diagnosis of bipolar depression is nearly exclusively made on the basis of a history of mania, independent of the cross-sectional psychopathological picture of the current depressive syndrome. We believe that it should be clinically evident that an operational definition of major depression should be wide enough to detect a rather high proportion of ICD-9 bipolar depressions. Measured by this goal none of the six operational definitions of major depression is able to offer a sufficiently high sensitivity: even RDC missed one-quarter of all ICD-9 bipolar depressions at the level of full diagnosis. The comparison of two extreme forms of wide and narrow definition of major depression demonstrates that only the wide definition (“meeting at least one of the six operational definitions of major depression”) is able to guarantee a very high (nearly 90%) sensitivity for the detection of ICD-9 bipolar depression.

The last two steps of applying time and exclusion criteria lead to the most important drop in sensitivity: the pure syndrome definition is still of excellent sensitivity (more than 90% for nearly all diagnoses). The reduction in sensitivity caused by the introduction of time criteria allows conclusions about the adequacy of different definitions of time criteria. FDC, RDC and I89 show the lowest reduction of sensitivity (between about 5% and 10%); these three diagnostic systems apply a liberal time criterion (only the total episode has to meet the time criterion). DSM, DSR and I87 require that every symptom has to meet the time criterion; these diagnostic systems lose between 22% and 26% sensitivity by applying this very strict time criterion. We conclude that these data clearly underline the greater adequacy of a more liberal time criterion, as used by FDC, RDC and I89. We can furthermore state that I89 is again superior to I87 and to DSM-III-R in choosing the more adequate time criterion.

Conclusions

The evaluation of concurrent validity using the independently assessed clinical ICD-9 diagnoses as a yardstick

has produced important results in the comparison of the competing operational definitions of major depression and for the adequacy of competing construction elements. Three essential conclusions can be drawn:

1. The "old" diagnostic definitions of RDC and FDC are superior to all newer definitions because they define the time criteria and the schizophrenic exclusion criteria more adequately than, for example, both DSM-III and DSM-III-R definitions.
2. The current ICD-10 definition I89 ("mild", "moderate" or "severe depression") comes closer to the concurrent validity of RDC and FDC than DSM-III, DSM-III-R and the previous ICD-10 I87 do.
3. All six competing diagnostic systems are too restrictive in identifying independently diagnosed ICD-9 bipolar depressions. Future developments of operational definitions of major depression should try to evaluate more liberal definitions to achieve at least 90% sensitivity for the detection of ICD-9 bipolar depressions.

However, these conclusions are restricted to only one aspect of concurrent validity; other validity aspects of in fact higher importance have still to be examined. Predictive validity will be the most important validation criterion under clinical aspects; criterion-related validity using family loading as a yardstick will be one of the most important validity aspects. Both dimensions of validity will be studied in forthcoming papers.

Acknowledgement. The study was conducted with the financial support of the Ministry of Health of the Federal Republic of Germany. We thank the following research assistants who performed part of the PODI ratings in the context of their doctoral theses: J. Böhm, M. Kancnyr, K. Lakatos, M. R. Lappas, C. Lesch, B. Wessels, S. Wustinger, M. Zwingerberger.

References

- American Psychiatric Association (1980) Diagnostic and Statistical Manual of Mental Disorders, 3rd edn (DSM-III). APA, Washington, W.C.
- American Psychiatric Association (1987) Diagnostic and Statistical Manual of Mental Disorders, 3rd edn, revised (DSM-III-R). APA, Washington, D.C.
- Cohen J (1960) A coefficient of agreement of nominal scales. *Educ Psychol Measurement* 20:37–46
- Delmo CD (1988) Weiterentwicklung des Polydiagnostischen Interviews. Medical dissertation. University of Mainz Medical School. Department of Psychiatry
- Delmo CD, Philipp M (1989) Computer program for calculating criteria, syndromes and diagnoses for the Polydiagnostic Interview PODI, version 2.1. Department of Psychiatry, University of Mainz
- Feighner JP (1981) Nosology of primary affective disorders and application to clinical research. *Acta Psychiatr Scand* 63 [Suppl 190]:29–41
- Feighner JP, Robins E, Guze SB, Woodruff RA, Winokur G, Munoz R (1972) Diagnostic criteria for use in psychiatric research. *Arch Gen Psychiatry* 26:57–63
- Hiller W, Mombour W, Rummeler R, Mittelhammer J (1988) Divergence and convergence of diagnoses for depression between ICD-9 and DSM-III-R. *Eur Arch Psychiatry Neurol Sci* 238:39–46
- Maier W, Philipp M, Buller R, Benkert O (1986) Sources of disagreement between clinical (ICD-9) and operational (DSM-III, RDC) diagnosis of endogenous depression (melancholia). *J Affect Dis* 11:235–243
- Philipp M, Maier W (1985) Operational diagnosis of endogenous depression. I. Comparison with clinical diagnosis. *Pharmacopsychiatry* 18:114–115
- Philipp M, Maier W (1986) The Polydiagnostic Interview: a structured interview for polydiagnostic classification of psychiatric patients. *Psychopathology* 19:175–185
- Philipp M, Maier W (1987) Diagnosesysteme endogener Depressionen. Springer, Berlin Heidelberg New York
- Philipp M, Maier W, Delmo CD (1990a) The concept of major depression. I. Descriptive comparison of six competing operational definitions including ICD-10 and DSM-III-R. *Eur Arch Psychiatry Neurol Sci* 240:229–236
- Philipp M, Maier W, Delmo CD (1990b) The concept of major depression. II. Empirical comparison of six competing operational definitions in 600 psychiatric inpatients. *Eur Arch Psychiatry Neurol Sci* 240:237–242
- Spitzer R, Endicott J, Robins E (1978) Research Diagnostic Criteria: rationale and reliability. *Arch Gen Psychiatry* 35:773–782
- World Health Organization (1987) April 1987 draft of the ICD-10 diagnostic criteria for research, chapter V: mental, behavioral, and developmental disorders (F00–F99). WHO, Geneva
- World Health Organization (1989) April 1989 draft of the ICD-10 diagnostic criteria for research, chapter V: mental, behavioral, and developmental disorders (F00–F99). WHO, Geneva